

19990106 030

Integrated Automatic Target Detection from Pixel-Registered Visual-Thermal-Range Images

Jorge E. Pérez-Jácome F. and Vijay K. Madisetti*

Center for Signal and Image Processing (CSIP), Georgia Institute of Technology - ECE,
Atlanta, GA 30332

Phone: (404)894-4696

e-mail: jorge@eedsp.gatech.edu, vkm@eedsp.gatech.edu

Abstract

This paper outlines a method to automatically detect targets from sets of pixel-registered visual, thermal, and range images. The method uses operations specifically designed for the different kinds of images. It also introduces the morphological operation called "erosion of strength n " as a powerful tool for removal of spurious information. Good preliminary results obtained for detection support its suitability for application to the Automatic Target Recognition (ATR) problem.

Introduction

By using multiple images from different sensors to detect and recognize targets, we can take advantage of the specific characteristics of the sensors and corresponding images and combine them to raise detection and/or recognition rates. A very good, yet brief, presentation of the different sensors used on ATR is given by Bhanu and Jones [1]. Our approach here is to define three basically different kinds of images according to what they represent, without concern for the specific method and/or sensor used to produce them:

Visual: Images that represent the intensity of the light emitted or reflected by bodies, within the visible band of the spectrum. A regular photograph is the typical example of this kind.

Thermal: Images whose pixel values represent a measure of the temperature at a specific location. Actually, they represent the intensity of light emitted or reflected by bodies, but inside a certain infrared region of the spectrum. Under certain conditions, the intensity obtained from an infrared [8–12 μ m] sensor is precisely related to the exact temperature by a

simple line equation [2].

Range: Images whose pixel values represent a measure of the distance from the objects to the sensor. On top-view aerial images, these images can represent elevation of terrain or objects.

The methods to produce the images can be very diverse: The sensors can be active or passive, they may use a given specific band or another. A given kind of image can be produced by different methods, but the resultant images are of the same nature, and so, can be operated on by algorithms defined for the specific kind of image. We now describe a method to perform target detection from sets of three pixel-registered images (visual, thermal, and range) for a given scene.

Detection Algorithm

Figure 1 presents the general scheme for the detection of targets from visual-thermal-range image sets. The system is designed to operate on top-view images with pixels represented by bytes (0 to 255). On the visual images, higher values represent brighter points. On the thermal images, higher values represent warmer points. The range images follow a format in which one-level increments correspond to changes of 10 cm in elevation. The resolution for the images is 25 cm per pixel, and the targets have rectangular to elliptical shapes, with an area of 150 to 2000 pixels. The different blocks of figure 1 are described next.

Bright/Dark point Extractor

The bright/dark-point extractor is used on both visual images and thermal images. It extracts points that are either darker or brighter than their surroundings in visual images, and points that are either warmer or colder than their surroundings in thermal images. Our method is similar to that of Nahm [3], with some modifications. Around a given pixel, it makes a rectangular annular window, one pixel wide, and estimates its mean $\mu_{i,j}$ and its standard deviation

*Through collaborative participation in the Advanced Sensors Consortium sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAL01-96-2001

UNCLASSIFIED

AQ I99-04-0515

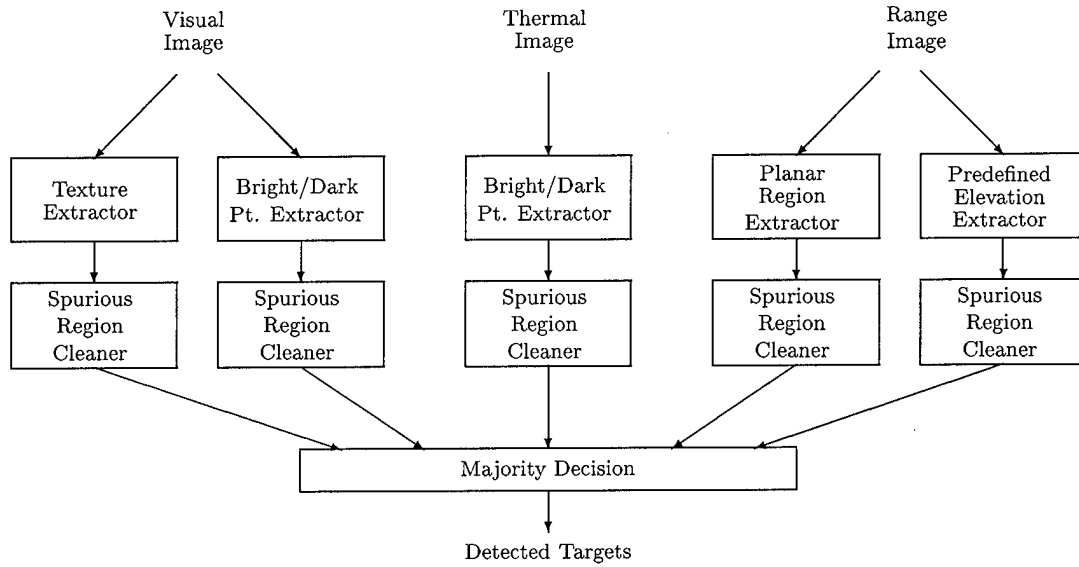


Figure 1: The detection algorithm

tion $\sigma_{i,j}$. Then, to determine the possibility that the pixel is part of a target, it checks whether its value $x_{i,j}$ differs from the average of the annular window by more than $1.5\sigma_{i,j}$. That is, if $(x_{i,j} - \mu_{i,j}) > 1.5\sigma_{i,j}$ or $(x_{i,j} - \mu_{i,j}) < -1.5\sigma_{i,j}$ then assign point (i, j) as a possible target.

Texture Extractor

The texture extractor operates on visual images. It measures the degree of similarity between adjacent pixels, for both the point under study (i, j) (presumably a target) and the pixels on an annular window around it (presumably clutter), and then compares them to see if they differ by more than a specified amount. As in the bright/dark-point extractor, we calculate a mean $\mu_{i,j}$ and a standard deviation $\sigma_{i,j}$ for the annular window, but of the absolute difference between adjacent pixels, rather than of their intensity. Also, we calculate $x_{i,j}$, the average difference between point (i, j) and its four adjacent points. So our test to determine a target point is similar to the one used for the the bright/dark-point extractor.

Planar Region Extractor

Since targets are well modeled by a collection of planar regions, the use of the degree of planarity to determine possible targets has been proposed [4]. A target usually has smooth (planar for small regions) surfaces, compared to most forms of clutters (grasses, trees, ground). The planar region extractor examines 3×3 pixel regions from the range images, and obtains an error e with respect to the equation of a plane

$z = ax + by + \rho_0$. Then it uses a threshold $e_{TH} = 0.6$, so a pixel is defined as a target if $e < e_{TH}$.

Predefined Elevation Extractor

If we have a basic knowledge of the kind of targets to search for (in our case, tanks), we can easily check if a point under study has an elevation suggesting a possible target. For this, we calculate $\mu_{i,j}$, the average elevation of a surface in an annular window around a point (i, j) , and then compare it with $x_{i,j}$, the elevation of the point (i, j) :

if $80\text{cm} < (x_{i,j} - \mu_{i,j}) < 250\text{ cm}$, then assign point (i, j) as a possible target.

Spurious Region Cleaner

After (or inside, when possible) the operators mentioned above, a downsampling of 4:1 is done, which reduces complexity, maintaining most of the detection information. Thus, we have smaller *binary* images (1: target, 0: no target), which have a series of regions that do not yet give a correct target detection. The problem is that there are many isolated single-pixel spots, or some spots that definitely do not have the shape of a target. Also, there are some spots that can be recognized by eye as targets, but which have many "holes" (pixels with value 0) in them. To remove this "noise," we use a series of morphological operations that are specifically designed for this purpose. We propose an erosion operator $eros-n(im, n)$, "erosion of strength n " which works as follows: a 3×3 template is passed over the binary image im . Around each pixel (i, j) , the number of 1's is counted. If it

Table 1: Detection results; (\checkmark) detected, (\times) miss.

Image	Extractor	Scene 1			Scene 2				Scene 3				SRC parameters	
		t_1	t_2	FA's	t_1	t_2	t_3	FA's	t_1	t_2	t_3	FA's	n_1	n_2
Visual	Texture	\checkmark	\times	9	\checkmark	\checkmark	\checkmark	28	\checkmark	\times	\times	16	3	5
Visual	Bright/Dark	\checkmark	\times	9	\checkmark	\times	\times	13	\checkmark	\times	\times	8	5	8
Thermal	Bright/Dark	\checkmark	\checkmark	1	\checkmark	\checkmark	\checkmark	3	\checkmark	\checkmark	\checkmark	5	7	8
Range	Pr. Elevation	\checkmark	\checkmark	6	\checkmark	\checkmark	\checkmark	10	\checkmark	\checkmark	\times	17	4	7
Range	Planarity	\checkmark	\checkmark	24	\checkmark	\times	\checkmark	21	\checkmark	\checkmark	\times	12	4	5
Overall Detection Results		\checkmark	\checkmark	1	\checkmark	\checkmark	\checkmark	2	\checkmark	\checkmark	\times	5	-----	-----

is larger than n , the output for the pixel (i, j) is 1, otherwise, it is 0. This operator can be used in series, with different values of n , and gives excellent results. This powerful but simple operator can be stated with MATLAB code as follows:

```
b=[1 1 1; 1 1 1; 1 1 1];
imOUT=conv2(imIN, b, 'same');
imOUT=(imOUT >= n);
```

When $n = 9$, it degenerates into the classical erosion operator. By using $n < 9$, we keep points on the input image that are important, but that would be eliminated with other erosion methods. Note also that the operator is independent of shape. An analogous "dilation of strength n " can also be defined. The spurious region cleaning (SRC) operator is defined as follows:

$$im_{out} = dilate(eros-n(eros-n(im_{in}, n_1), n_2))$$

The application of two erosion operators in series results in great performance on eliminating spurious target pixels, for different densities and target-to-clutter contrasts. The parameters n_1 and n_2 are toned to specifically work on the different images. The *dilate* operator is used to join together points that are likely to belong to the same target. The output gives small regions inside the likely targets, usually with very few false alarms, for the different binary images.

Majority Decision

The final function of the detector is to combine the results of the individual detectors to produce the final output. The method we use checks the five detectors, and for each pixel it assigns a 1 if there are three or more 1s as inputs, and assigns a 0 otherwise. If a cluster of 1s overlaps a target, the target is declared *detected*, otherwise it is declared a *miss*. A cluster not overlapping a target is declared a *false alarm* (FA).

Test Data and Results

We analyzed three different sets of images (each consisting of a visual, a thermal, and a range image),

representing three scenes. The first scene has two tanks, on a dry area, without vegetation. The second scene has three tanks, including one partially occluded by vegetation. The third scene has also three tanks including one partially occluded by vegetation, and it has several pieces of cultural clutter, such as small buildings, bridges, etc. The last two scenes have bodies of water as well. The images, 512×512 pixels, are artificial, but were synthesized with information from real visual images.

The generation process was as follows: Visual backgrounds were taken from selected aerial photographs. These images were clipped and scaled to match our objectives. Then, we embedded visual images of tanks on the images, with the use of interactive programs. The location and orientation of the targets were chosen to resemble a real scene as closely as possible. Then thermal images were first generated with the use of interactive tools to define temperature values for every part of the images, and then post-processed with the use of filtering, interpolation, and the addition of spatially correlated random data. Finally, the range images were synthesized in a similar way, incorporating not only the elevation data, but the random height variability of the different surfaces that composed the scene, as would result from sub-pixel information.

As seen in table 1, some targets could not be detected from individual images, but most were correctly detected by our integrated system. Of the total of eight targets, seven were correctly detected, and only eight false alarms were produced, most of them from cultural clutter in the third scene. The number of false alarms per scene was at most as large as for the best individual detector. The only *miss* corresponds to a partially occluded target. We believe that our system would perform even better on noisier thermal images. We plan to apply our system to additional real scenes when data becomes available, to see its actual perfor-

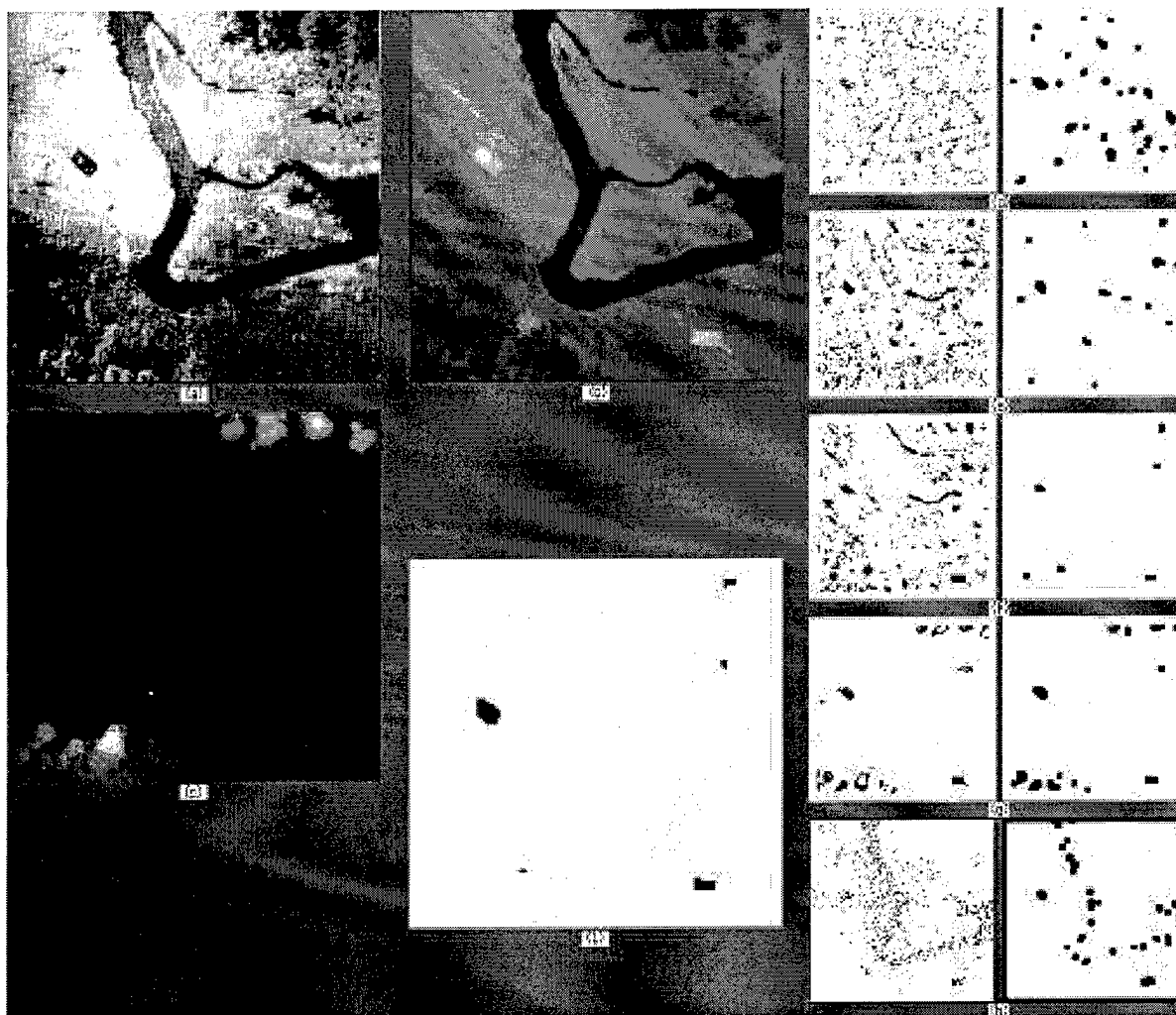


Figure 2: Detection process for scene 2. (a-c) Original visual, thermal, and range images. (d-h) show the detection before SRC (left) and after SRC (right), as they result from: (d) texture extractor on visual image, (e) bright/dark extractor on visual image, (f) bright/dark extractor on thermal image, (g) predefined elevation extractor on range image, (h) planar region extractor on range image. (i) overall detection results.

mance and to adjust the parameters correspondingly. We also think that the process of recognition can be greatly improved by the use of this scheme. Once the targets are detected, it is easier to perform direct template matching for the different kinds of images. A series of images with the complete detection results for scene 2 is shown in figure 2; t_1 , t_2 and t_3 are in the left, bottom left, and bottom right respectively, t_2 is partially occluded by foliage.

The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies either expressed or implied, of the Army Research Laboratory or the US Government.

References

- [1] B. Bhanu and T. L. Jones, "Image Understanding Research For Automatic Target Recognition," IEEE Aerospace and Electronic Systems Magazine, Vol. 8, No. 10, pp. 15-23, October, 1993.
- [2] Nandhakumar and Aggarwal, "Integrated Analysis of Thermal and Visual Images for Scene Interpretation," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 10, No. 4, July, 1988.
- [3] J. W. Nahm, "Joint ATR-Compression for FLIR and SAR Images," proposal for doctoral dissertation, Georgia Tech, 1996.
- [4] Roggemann, Mills, Kabrisky, Rogers, Tatman, "An Approach to Multiple Sensor Target Detection," SPIE Sensor Fusion II, Vol. 1100, pp. 42-52, 1989.

INTERNET DOCUMENT INFORMATION FORM

A . Report Title: Integrated Automatic Target Detection from Pixel-Registered Visual-Thermal-Range Images

B. DATE Report Downloaded From the Internet 1/5/99

C. Report's Point of Contact: (Name, Organization, Address, Office Symbol, & Ph #): Georgia Institute of Technology-ECE
Jorge E. Perez (404) 894-4696
Center for Signal and Image
Processing (CSIP)
Atlanta, GA 30332

D. Currently Applicable Classification Level: Unclassified

E. Distribution Statement A: Approved for Public Release

F. The foregoing information was compiled and provided by:
DTIC-OCA, Initials: VM_ **Preparation Date:** 1/5/99__

The foregoing information should exactly correspond to the Title, Report Number, and the Date on the accompanying report document. If there are mismatches, or other questions, contact the above OCA Representative for resolution.